

See discussions, stats, and author profiles for this publication at: [https://www.researchgate.net/publication/309308779](https://www.researchgate.net/publication/309308779_The_effect_of_zonal_factors_in_estimating_crash_risks_by_transportation_modes_Motor_vehicle_bicycle_and_pedestrian?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_2&amp;_esc=publicationCoverPdf)

[The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian](https://www.researchgate.net/publication/309308779_The_effect_of_zonal_factors_in_estimating_crash_risks_by_transportation_modes_Motor_vehicle_bicycle_and_pedestrian?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_3&amp;_esc=publicationCoverPdf)

**Article** *in* Accident Analysis & Prevention · January 2017

DOI: 10.1016/j.aap.2016.10.018

CITATIONS

8

READS

119

**3 authors**, including:

[Helai Huang](https://www.researchgate.net/profile/Helai_Huang?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_5&amp;_esc=publicationCoverPdf)

[University of Texas at Arlington](https://www.researchgate.net/institution/University_of_Texas_at_Arlington?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_6&amp;_esc=publicationCoverPdf)

**72** PUBLICATIONS **1,588** CITATIONS

[Qiang Zeng](https://www.researchgate.net/profile/Qiang_Zeng10?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_5&amp;_esc=publicationCoverPdf)

[South China University of Technology](https://www.researchgate.net/institution/South_China_University_of_Technology?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_6&amp;_esc=publicationCoverPdf)

**23** PUBLICATIONS **184** CITATIONS

# 

[SEE PROFILE](https://www.researchgate.net/profile/Helai_Huang?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_7&amp;_esc=publicationCoverPdf)

[SEE PROFILE](https://www.researchgate.net/profile/Qiang_Zeng10?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_7&amp;_esc=publicationCoverPdf)

All content following this page was uploaded by [Qiang Zeng](https://www.researchgate.net/profile/Qiang_Zeng10?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_10&amp;_esc=publicationCoverPdf) on 10 October 2017.

The user has requested enhancement of the downloaded file.

# 1

1. **Research highlights**
2.  This paper investigates factors associated with crash occurrences by different
3. transportation modes.
4.  Zonal factors have significant effects on intersection crashes.

#  Zonal factors contribute to tracking the source of heterogeneity of risk factors.

1.  Zonal factors play a more important role in elevating the model performance
2. for non-motorized than motor-vehicle crashes.
3.  A relatively smaller buffer width to extract zonal factors yields better
4. estimations.

11

12

1. **The Effect of Zonal Factors in Estimating Crash Risks by Transportation**
2. **Modes: Motor Vehicle, Bicycle and Pedestrian**

3

4 Jie WANG1, Helai HUANG1\*, Qiang ZENG2

# 5

1. 1 School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China
2. 2 School of Civil Engineering and Transportation, South China University of Technology, Guangzhou,
3. Guangdong, China
4. \*Correspondence
5. E-mail address: [jie\_wang@csu.edu.cn](mailto:jie_wang@csu.edu.cn) (J. Wang), [huanghelai@csu.edu.cn](mailto:huanghelai@csu.edu.cn) (H. Huang)**,**
6. [zengqiang@scut.edu.cn](mailto:zengqiang@scut.edu.cn) (Q. Zeng)

# 12

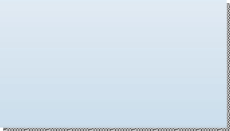
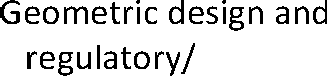
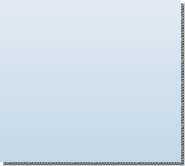
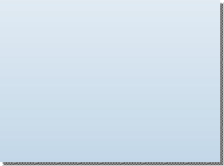
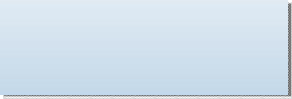
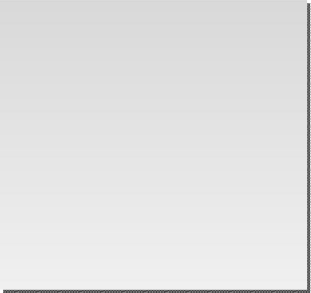
1. **Abstract:** *Objectives:* This paper aimed to (i) differentiate the effects of contributory factors on
2. crash risks related to different transportation modes, i.e., motor vehicle, bicycle and pedestrian; (ii)
3. explore the potential contribution of zone-level factors which are traditionally excluded or
4. omitted, so as to track the source of heterogeneous effects of certain risk factors in
5. crash-frequency models by different modes. *Methods:* Two analytical methods, i.e. negative
6. binomial models (NB) and random parameters negative binomial models (RPNB), were
7. employed to relate crash frequencies of different transportation modes to a variety of risk factors
8. at intersections. *F*ive years of crash data, traffic volume, geometric design as well as macroscopic
9. variables at traffic analysis zone (TAZ) level for 279 intersections were used for analysis as a case
10. study. *Results:* Among the findings are: (1) the sets of significant variables in crash-frequency
11. analysis differed for different transportation modes; (2) omission of macroscopic variables would
12. result in biased parameters estimation and incorrect inferences; (3) the zonal factors (macroscopic
13. factors) considered played a more important role in elevating the model performance for
14. non-motorized than motor-vehicle crashes; (4) a relatively smaller buffer width to extract
15. macroscopic factors surrounding the intersection yielded better estimations.
16. **Keywords:** transportation modes; macroscopic variables; unobserved heterogeneity; buffer
17. width; intersection safety

# 30

## 1 Introduction

1. Many communities have increased their interest in the implementation of multimodal
2. transportation and advocated for the shift from motor vehicles to non-motorized modes of
3. transportation, i.e., walking and cycling. In spite of the health and environmental benefits, an
4. increasing number of crashes involving pedestrians and bicyclists has become a major concern in
5. improving traffic safety. For example in 2013, the United States had 4735 pedestrian and 743
6. bicyclist deaths, accounting for 18% of all U.S. highway fatalities (NHTSA, 2013). The Federal
7. Highway Administration’s office of safety has established pedestrian and bicyclist safety as one
8. of its top priorities. Thus, it is essential for traffic safety engineers to provide appropriate
9. countermeasures or policies to achieve friendly and safe multimodal transportation.
10. A comprehensive understanding of contributing factors associated with crash occurrences by
11. different modes is a prerequisite for developing safety improvement programs to effectively
12. reduce traffic crashes. For a given road entity (e.g. road segments or intersections), the potential
13. factors associated with multimodal crashes could be summarized as in Figure 1, according with
14. Miranda-Moreno et al. (2011), Mitra and Washington (2012), Ukkusuri et al. (2012) and Strauss et
15. al. (2003; 2014). The factors influencing road-entity-level crash frequency by modes include
16. macroscopic factors related to built environment of the road entities - such as population and
17. economic characteristics, land use characteristics and travel behaviors - as well as road features
18. and traffic characteristics of the road entities. In addition, crash occurrence is also associated with
19. individual characteristics such as gender, age, education, alcohol consumption, and other driver
20. and pedestrian behaviors (Ryb et al., 2007). Although discrete individual-level factors are not
21. available to be integrated into the crash-frequency model, individual characteristics are always
22. influenced by macroscopic factors (Christoffel and Gallagher, 1999). Therefore, macroscopic
23. factors could serve as a surrogate for individual behaviors.

# 7



|  |
| --- |
|  |
|  |
|  |

8

9 Figure 1. Factors associated with multimodal crashes

# 10

1. The choice of appropriate analytical method and the selection of representative explanatory
2. variables are two important considerations for obtaining accurate model predictions. Over the
3. past three decades, considerable research efforts have been devoted to developing and applying
4. sophisticated methodological approaches associated with the analysis of crash frequency.
5. Detailed descriptions and assessments of crash-frequency models can be found in the review
6. papers by Lord and Mannering (2010) and Mannering and Bhat (2014). However, relatively few
7. studies have focused on the identification and inclusion of traditionally excluded or omitted
8. variables in crash-frequency analysis. In particular, variables related to macroscopic factors
9. previously described (in Figure 1) are normally unavailable in crash databases and as a result
10. have rarely been examined in great detail. Mitra and Washington (2012) is one of a few studies
11. exploring the omitted variables in crash-frequency modeling. The authors developed two
12. different models of estimating intersection crash frequency, one with traffic volume as the only
13. independent variable, and the other with several spatial factors in addition to commonly
14. included geometric design and traffic factors. Through contrastive analysis of the two models,
15. results indicated that some spatial factors, such as local influences of weather, sun glare,
16. proximity to drinking establishment, proximity to school and demographic attributes near
17. intersections, have significant explanatory power and their exclusion leads to biased estimates.
18. Statistical methods such as spatial and temporal correlation, multilevel, random effect,
19. random parameter, and latent class approaches have been developed to address this issue of
20. unobserved heterogeneity (Anastasopoulos and Mannering, 2009; Dong et al., 2016; Mannering et
21. al., 2016; Quddus, 2008; Wang and Huang, 2016; Xu and Huang, 2015; Xu et al. 2016), as these
22. omitted explanatory variables can be regarded as part of the unobserved heterogeneity.
23. Unobserved heterogeneity impacts traffic safety analysis in two ways: the first problem is that the
24. selected explanatory variables cannot fully account for the cross-section or longitudinal-section
25. variations in crash counts due to unobserved road geometrics, environmental factors, driver
26. behavior and other confounding factors, which lead to impaired predictive performance of the
27. model (called heterogeneity in model prediction); the second problem is that these unobserved
28. factors are always correlated with observed factors and thus biased parameters will be estimated
29. and incorrect inferences could be drawn (called heterogeneity in the coefficient estimator). While
30. these approaches will mitigate the adverse impacts of omitting significant explanatory variables,
31. the resulting model estimates still fail to track the original source of heterogeneity and quantify
32. the safety effect of omitted variables (such as macroscopic factors shown in Figure 1). Omission of
33. important explanatory variables still remains a problem even with advanced statistical
34. approaches to capture unobserved heterogeneity (Mannering et al., 2016).
35. The study by Mitra and Washington (2012) attempted to investigate the safety effect of some
36. important omitted variables on total crash frequency and their contribution on model estimation.
37. As Venkataraman et al. (2013) stated, frequency models of crash outcome type can provide
38. substantial insights into the effect of explanatory variables and assist in examining the
39. heterogeneity effects in roadway geometric features. This paper aims to extend previous research
40. (Mitra and Washington, 2012) and investigate how macroscopic factors affect the crash-frequency
41. analysis for different transportation modes. This is because there may be some inconsistent
42. impacts of some macroscopic variables on motor vehicle and non-motorized (including bicycle
43. and pedestrian) crashes. For example, Lee et al. (2015) utilized a multivariate model for
44. investigating motor vehicle and non-motorized crashes at the macroscopic level. Results for the
45. parameter estimation suggested that some zonal variables related to demographics and road
46. characteristics have different directional effects on motor vehicle and non-motorized crashes.
47. Meanwhile, the most appropriate width of buffer to extract macroscopic factors may be
48. inconsistent between modeling motor vehicle and non-motorized crashes. Therefore, it is
49. advisable to model the crash frequency by separate transportation modes to examine the effects
50. of macroscopic factors.
51. In summary the objective of this paper is twofold: (1) to examine the effects of a host of
52. contributing factors including both macroscopic and microscopic factors on crash occurrence
53. with respect to different transportation modes; (2) to shed further light on the contribution of the
54. macroscopic factors which are traditionally excluded or omitted variables, to tracking the source
55. of heterogeneity effects in coefficient estimators of regularly used variables and improving the
56. model performance in crash-frequency analysis related to different modes.

## 2 Data preparation

1. In this study, data collected for 279 intersections located in Hillsborough County, Florida,
2. USA were used to develop the intersection crash-frequency models for different transportation
3. modes. The data for the analysis was mainly divided into four types: traffic crash data, traffic
4. characteristics, road characteristics related to geometric design, traffic control/regulatory of the
5. intersection, macroscopic factors including trip production/attraction, demographic and
6. socio-economic characteristics surrounding the intersection. The derivation and processing of
7. these data sources are described next.
8. *2.1 Crash data*
9. Crash data for the intersections in a five-year period (2005–2009) were obtained from the
10. Florida Department of Transportation (FDOT) Crash Analysis Reporting (CAR) system. Crashes
11. were categorized as intersection related crashes if they occurred within the curb-line limits of
12. the intersection or if they occurred within the influence area of the intersection, which is 250 feet
13. away from the stop line. Intersection-level crash data was disaggregated into motor vehicle,
14. bicycle, and pedestrian crashes. A motor vehicle crash was defined as a collision between two or
15. more motor vehicles or between a motor vehicle and an object. A bicycle crash referred to a
16. collision between a motor vehicle and a bicycle. Likewise, a pedestrian crash denoted a collision
17. between a motor vehicle and a pedestrian.
18. *2.2 Traffic characteristics*
19. Previous researches suggest that traffic characteristics such as motor vehicle, pedestrian and
20. bicycle volume are the most important factors influencing crash occurrences. Motor vehicle
21. volume represented by average annual daily traffic (AADT) can be collected from the FDOT
22. Roadway Characteristics Inventory. Two motor vehicle volume variables including AADT from
23. major road and AADT from minor road of 5-year (2005–2009) average were also obtained. Actual
24. pedestrian and bicycle volume are not regularly available. The collected macroscopic data such as
25. population were used to serve as a surrogate for pedestrian and bicycle volume as suggested by
26. Jacobsen (2003) and Miranda-Moreno et al.(2011).
27. *2.3 Road features*
28. Road features related to geometric design and regulatory/control attributes of the road entity
29. were collected from the FDOT Roadway Characteristics Inventory. The road factors considered in
30. the study are number of legs, presence of traffic signal, speed limit on major approach, and speed
31. limit on minor approach.
32. *2.4 Macroscopic factors*
33. Considerable previous studies on zonal-level crash-frequency models suggests that various
34. macroscopic factors such as trip production/attraction, demographic and socio-economic
35. characteristics affect area-wide traffic crashes (Quddus., 2008; Huang et al., 2010, 2016; Abdel-Aty
36. et al., 2011; Xu et al., 2014; Dong et al., 2015, 2016). It is hypothesized that the number of crashes
37. occurring at an intersection is also associated with these macroscopic factors surrounding the
38. intersection.
39. Trip production/attraction factors such as total trip productions/attraction, home-based work
40. productions/attraction, college productions/attraction at the TAZ level were collected from the
41. Intermodal Systems Development Unit of District 7 of the FDOT. Demographic and
42. socio-economic characteristics were examined including the geographical area of each TAZ,
43. population, income and commuting, which were downloaded from the United States Census
44. report.
45. ArcGIS 10.0 was used to generate a buffer around each selected intersection and conduct a
46. spatial analysis to extract macroscopic data from the TAZ layers. The process is described in
47. detail as follows. First, a spatial overlay of TAZ layers on a specified width buffer (including
48. 0.25 mile, 0.5 mile and 1 mile buffer) was generated around each intersection. Then, spatial
49. analysis operators such as “intersect” and “join” available in the GIS environment were used to
50. intersect layers, join tables and extract selected trip production/attraction, demographic and
51. socio-economic characteristics within the generated buffer. Macroscopic factors were distributed
52. in proportion to the area of TAZ within the generated buffer. The ArcGIS procedure adopted to
53. extract and estimate macroscopic factors in this study was similar to the one discussed in detail
54. by Pulugurtha and Sambhara (2011) to develop pedestrian crash estimated models.
55. Table 1 provides descriptive statistics of crash data, traffic variables, road variables and
56. macroscopic variables located in 0.5 mile buffer. The values of macroscopic variables located at
57. 0.25 and 1 mile buffer are not listed for compactness of the table.

# 38

39

40

41

42

43

44

45

46

47

1. Table 1 Summary of variable and descriptive statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Definition** | **Meana** | **SDa** | **Mina** | **Maxa** |
| **Crash data** |  |  |  |  |  |
| Motor vehicle crash | Motor vehicle crash per intersection in 2005-2009 | 65.219 | 56.545 | 2.000 | 293.00 |
| Bicycle crash | Bicycle crash per intersection in 2005-2009 | 1.018 | 1.340 | 0.000 | 8.000 |
| Pedestrian crash | Pedestrian crash per intersection in 2005-2009 | 1.276 | 1.889 | 0.000 | 11.000 |
| **Traffic and road variables** | | | | | |
| AADT-major | AADT on major approach (103pcu) | 28.364 | 17.684 | 2.600 | 71.300 |
| AADT-minor | AADT on minor approach (103pcu) | 9.150 | 8.879 | 1.000 | 43.000 |
| Leg-number | Number of legs (4 legs=1, 3 legs=0) | 0.670 | 0.471 | 0.000 | 1.000 |
| traffic signal | Presence of traffic signal (yes=1,no=0) | 0.498 | 0.542 | 0.000 | 4.000 |
| Speed-Major | Speed limit on major approach (mph) | 40.502 | 6.154 | 6.154 | 60.000 |
| Speed-Minor | Speed limit on minor approach (mph) | 35.323 | 6.479 | 6.479 | 55.000 |
| **Macroscopic variables** |  |  |  |  |  |
| PA\_density | Density of productions and attractions (per acre) | 49.272 | 27.369 | 2.514 | 184.06 |
| HB\_prop | Proportion of home-based productions and attractions | 0.680 | 0.084 | 0.317 | 0.840 |
| Col\_prop | Proportion of college productions and attractions | 0.025 | 0.059 | 0.000 | 0.415 |
| Pop\_density | Density of total population (per acre) | 5.631 | 2.554 | 0.342 | 11.798 |
| Age 0 to 15\_ prop | Proportion of population between age 0 and 15 | 0.225 | 0.050 | 0.069 | 0.348 |
| Age 16 to 64\_prop | Proportion of population between age 16 and 64 | 0.652 | 0.054 | 0.561 | 0.917 |
| Pub\_prop | Proportion of workers commuting by public transportation | 0.030 | 0.031 | 0.000 | 0.149 |
| Wal\_ prop | Proportion of workers commuting by walking | 0.025 | 0.017 | 0.000 | 0.080 |
| MHINC | Median household income (in thousands) | 36.728 | 15.199 | 4.300 | 89.035 |

1. a These values relating to macroscopic variables are only for 0.5 mile buffer.

## 3 Methodology

1. In previous crash-frequency analyses, Poisson and Negative binomial model (NB), along
2. with their variants (such as Poisson-lognormal model), are commonly used and proven to be
3. successful as they effectively model the rare, random, sporadic, and non-negative crash data. As
4. crash data exhibit over-dispersion (i.e., variance greater than mean), NB is superior to the Poisson
5. model. Compared with the basic Poisson model, NB includes a gamma-distributed error term in
6. Poisson mean to account for the over-dispersion due to omission of relevant variables or
7. measurement error in crash data. The formulation for NB can be presented as follows:

11

1. *Yi* ~ Poisson*i* 

ln *i*  **0  **X***i* **β**  * i*

(1 )

(2 )

1. Where *Yi* is the crash frequency by modes (i.e., motor vehicle, bicycle and pedestrian) at
2. intersection *i* , and *i* is the expectation of *Yi* . **X***i* is a vector of explanatory variables. **0

is the

1. intercept, **β** is a vector of estimable parameters. *i* is the error term that is assumed to be
2. independent **X** and has a two-parameter gamma distribution.
3. The NB model presented in Eq.(2) could control for unobserved heterogeneity by omitted
4. variables. However, this model assumes that the unobserved variables are uncorrelated with the
5. observed exploratory variables. If this correlation exists, unobserved factors can introduce
6. variation in the effect of observed variables on crash likelihood. Random parameters approaches
7. are able to address this issue by allowing non-constant estimable parameters to vary across
8. observations (Mannering et al., 2016). In random parameters negative binomial model (RPNB),
9. estimable parameters ( **β** ) in Eq.(2) can be written as:

3

*i*  **  *i*

1. Where ** is the mean of the random parameter

*i* ,

(3 )

*i* is a randomly distributed term (e.g.,

1. a normally distributed term with mean 0 and variance ** 2 ) that capture heterogeneity across

**

1. observations. The analyst can test for random parameters with Eq.(3), across all observations i for
2. each included explanatory variable. If the variance of the chosen distribution is not significantly
3. different from zero, it suggests that a conventional fixed parameter is statistically appropriate.
4. Thus the model always combines fixed and random parameters across the included explanatory
5. variables.
6. As previously stated, heterogeneity effects of certain risk factors mainly derive from the
7. combined effects of unobserved variables that have been omitted from the model. Although the
8. random parameters approaches could mitigate the adverse impacts of omitting variables, the
9. original source of unobserved heterogeneity (or what are the major factors that lead to
10. unobserved heterogeneity) still fails to be well understood. Thus one of the aims of this study is
11. to test the potential role of macroscopic variables (referring to as the influential omitted variables)
12. in tracking the source of heterogeneity effects related to commonly used traffic and road
13. variables.
14. To this end, two different model specifications are estimated and compared, both with
15. random parameters approaches. In the base model only traffic and road variables are included,
16. while the second model traffic and road variables as well as macroscopic variables at TAZ level
17. surrounding the intersections are included. If the test results for parameter estimation on a
18. variable appear as random in base model (variance of the chosen distribution is significantly
19. different from zero), while this variable has the fixed effect in the second model. Then we could
20. infer that the heterogeneity effect of this variable is mainly caused by these macroscopic variables.
21. For another case, the safety effect of this variable is still random in second model; the source of
22. heterogeneity effect of this variable still cannot be clearly distinguished, maybe due to other
23. important omitted variables.
24. Apart from the potential role in accounting for the heterogeneity effects in parameters
25. estimation, integrating the macroscopic variables could also decrease the variance of the random
26. error (i.e. overdispersion) and thus improve the model performance in crash-frequency
27. prediction. The proportion of reduction in variance (PRV), also called explained variance,
28. proposed by Raudenbush and Bryk (2002) can be used to assess the overall explanatory power of
29. macroscopic factors for modeling the crashes by different modes. In this case, the PRV of
30. macroscopic variables is defined as:

36

** 2 - ** 2



*PRV*

0 1

2

**

0

**

(4 )

37 Where

**

2

0 is the variance of the error term in the base model without macroscopic variables.

38 2

1

is the variance of error term in the full model with macroscopic variables. The value of PRV is

1. bounded by 0 and 1, and a higher value indicates a stronger explanatory power of macroscopic
2. factors on the crash occurrence.
3. Furthermore, two goodness-of-fit statistics are used for model comparisons: Akaike
4. Information Criterion (AIC) and log-likelihood ratio (LR).
5. The AIC is calculated as follows:

44

*AIC*  2*LL*  2 *p*

(5)

1. Where LL is the log-likelihoods at convergence for the estimated model, and *p* is the
2. number of parameters in the statistical model. The model with the lower AIC is considered to
3. have the better goodness of fit.
4. The LR value is the chi-squared value in the log-likelihood ratio test for the null hypothesis
5. test that reveals whether or not the equivalence of two models should be rejected. The likelihood
6. ratio statistic is,

5

*LR*  2(*LLN*  *LLA* )

(6)

1. which is ** 2 distributed with J degrees of freedom, where J = KA − KN (KA and KN are the
2. number of coefficients for the alternative model and the null model, respectively ), LLN and LLA are
3. the log-likelihoods at convergence for the null model and the alternative model, respectively. The
4. null hypothesis for Eq. (6) is that the alternative model does not have a significantly lower
5. log-likelihood than the null models, indicating a lack of significant difference between the null
6. model and the alternative model.

## 4 Results and discussion

1. Three types of crash-frequency models for motor vehicles, bicycles and pedestrians were
2. developed. Each type of model involved eight separate models based on four model
3. specifications (one with only traffic volume and road features, the other three with macroscopic
4. factors overlaid on 0.25, 0.5, 1 mile buffer respectively in addition to commonly included traffic
5. volume and road features ) and two analytical methods (NB and RPNB).
6. LIMDEP econometric software was used to develop the statistical models described above.
7. To enable focus on the most significant variables, variables that were not found to significantly
8. different from zero at the 0.1 level of significant using a *t*-test were removed. Meanwhile, the
9. likelihood ratio test was used to guarantee that each added variable significantly improved the
10. overall model performance. In the RPNB, if the variance of a random parameter was not
11. statistically different from zero, the random parameter was simplified to be fixed across
12. intersections. Thus, the results in NB were in accordance with that in RPBN when no estimate
13. parameters of explanatory variable were statistically random.
14. This analysis below will emphasize testing effects of macroscopic factors on model
15. performance in crash-frequency analysis for three transportation modes, and then comparison
16. results of the parameter estimates and marginal effects between the base model and the full
17. model with macroscopic variables will be presented and interpreted.
18. *4.1 Effects of macroscopic factors on model performance*
19. Tables 2-4 show goodness-of-fit measures for motor vehicle, bicycle and pedestrian
20. crash-frequency models, respectively. As shown in Table 2, only five of eight motor vehicle
21. crash-frequency models, including two base models (NB model and RANB model) and three
22. fully specified NB models, were presented since there was no significant random parameter as
23. measured by the t-statistics in all three models with macroscopic variables. Although there was
24. no substantial difference in goodness-of-fit as reflected by likelihood ratio test between the base
25. NB and RPNB model, the present of significant random parameters ( e.g. the variable of ‘presence
26. of traffic signal’ in this case study) demonstrated the existent of heterogeneity of risk factors in
27. base model without considering macroscopic factors. More interestingly, no significant random
28. parameters were found in all three full models with macroscopic variables. This implied that the
29. heterogeneous effects of risk factors on motor vehicle crash frequency could be mostly captured
30. by these macroscopic variables, at least for the Hillsborough dataset examined here. Frequency
31. analysis models for bicycle crashes presented a similar result to motor vehicles crashes, as shown
32. in Table 3. However, this was not the case for the pedestrian crash-frequency models (Table 4).
33. Significant random parameters, such as ‘presence of traffic signal’, existed both in pedestrian
34. crash models with and without macroscopic variables, suggesting that the heterogeneity effect in
35. parameters estimation cannot be completely picked up by these macroscopic variables.
36. Apart from the potential effect in tracking the heterogeneity, results also revealed that
37. incorporating the macroscopic variables in crash-frequency analysis leaded to an increasing
38. model complexity but a considerable improvement in overall fit as measured by log likelihood at
39. convergence. As shown in Table2, the likelihood ratio test comparing the full NB models and the
40. base NB models indicated that we were more than 99.99 % confident that the full models with
41. macroscopic variables (except the full model with 1.0 mile buffer-width macroscopic variables)
42. were statistically superior. This comparison suggested that the macroscopic variables explained a
43. portion of variability in crash occurrences and should not be omitted in motor vehicle
44. crash-frequency model. In regard to bicycle and pedestrian models, omission of macroscopic
45. variables will also lead to a significant decrease in goodness-of-fit, as shown in Tables 3-4.
46. Comparing model outputs developed based on 0.25 mile, 0.5 mile and 1 mile buffer width
47. data, these models with macroscopic variables of 0.25 mile buffer width had the lowest AIC,
48. conversely, the models with macroscopic variables of 1 mile buffer width had the highest AIC in
49. all three types of crash-frequency models by modes. Thus, a relatively smaller buffer width in
50. extracting macroscopic factors surrounding the intersection would yield a better estimate.
51. To further assess and compare the overall explanatory power of macroscopic variables, the
52. values of PRVs were calculated. As shown in Table 2, the motor vehicle crash-frequency model
53. with macroscopic variables of 0.25 mile buffer width had the highest PRV of 7.98%. This meant
54. that 7.98% of unexplained variation resulted from those omitted macroscopic variables, which
55. also suggested the usefulness of the motor vehicle crash-frequency analysis by integrating
56. macroscopic factors. Accordingly, the highest values of PRV were 33.02% and 26.37% in bicycle
57. and pedestrian crash-frequency models respectively, as shown in Tables 3-4.
58. Table 2 Goodness-of-fit measures for motor vehicle crash-frequency models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model statistics | Base | model | 0.25 mile | 0.50 mile | 1 mile |
|  | NB | RPNB | NB | NB | NB |
| Number of observers | 279 | 279 | 279 | 279 | 279 |
| Number of parameters | 6 | 7 | 10 | 10 | 10 |
| Log likelihood at convergence | -1280.07 | -1279.42 | -1269.25 | -1270.86 | -1276.28 |
| AIC | 2572.14 | 2572.84 | 2558.50 | 2561.73 | 2572.56 |
| Log-likelihood ratio test | | | | | |
| 2 = -2(LLN-LLA) | 1.300 | | 21.645 | 18.416 | 7.588 |
| Degrees of freedom | 1 | | 4 | 4 | 4 |
| P-value | 0.26 | | <0.01 | <0.01 | 0.11 |
| Explanatory power of macroscopic factors | | | | | |
| Variance of the error term, | 0.250 | 0.246 | 0.230 | 0.232 | 0.243 |
| Proportion of reduction in variance, PRV |  | 1.23% | 7.98% | 6.85% | 2.67% |

1. Note: LLN denotes the log likelihood at convergence for Base + NB model.

# 26

1. By comparing PRVs in models by different transportation modes, the PRVs in bicycle and
2. pedestrian crash-frequency model were much higher than that in motor vehicle models. In other
3. words, integrating macroscopic factors in non-motorized crash-frequency model was more vital
4. than that in developing motor vehicle model. This result was in line with the expectation. One
5. possible reason for this distinct effect is that pedestrian/bicycle volume (or pedestrian/bicycle
6. activity) which is commonly identified as the main determinants of pedestrian/ bicycle crash
7. frequency has been omitted in the base pedestrian/bicycle crash-frequency models. Integrating
8. macroscopic factors for pedestrian/bicycle crash-frequency analysis made up the absence of
9. pedestrian/bicycle volume in predicting pedestrian/bicycle crash frequencies to some extent as
10. demonstrated by previous study (Jacobsen, 2003; Miranda-Moreno et al., 2011) that macroscopic
11. data can serve as a surrogate for pedestrian and bicycle volume. Another reason, maybe even
12. more importantly, originates from the differences in the travel distance between non-motorized
13. and motor vehicle modes. As walking and bicycle are short-distance transportation modes, crash
14. victims of pedestrians and bicyclists generally reside near the crash intersection, and thus, the
15. macroscopic factors extracted surrounding the intersection can probably better reflect pedestrian
16. and/or bicyclist behaviors than that for motor drivers.
17. Table 3 Goodness-of-fit measures for bicycle crash-frequency models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model statistics | Base  NB | model  RPNB | 0.25 mile  NB | 0.50 mile  NB | 1 mile  NB |
| Number of observers | 279 | 279 | 279 | 279 | 279 |
| Number of parameters | 6 | 7 | 8 | 8 | 8 |
| Log likelihood at convergence | -362.53 | -361.61 | -350.42 | -351.14 | -351.62 |
| AIC | 737.06 | 737.22 | 716.85 | 718.28 | 719.24 |
| 2 = -2(LLN-LLA) | Log-likelihood ratio test  1.837 | | 24.209 | 22.777 | 21.821 |
| Degrees of freedom | 1 | | 2 | 2 | 2 |
| P-value | 0.17 | | <0.01 | <0.01 | <0.01 |
| Explanatory power of macroscopic factors  Variance of the error term 0.402 0.335 0.269 | | | | 0.274 | 0.280 |
| Proportion of reduction in variance, PRV 16.48% 33.02% | | | | 31.66% | 30.27% |
| 10 Note: LLN denotes the log likelihood at convergence for Base + NB model.  11 | | | |  |  |
| 12 Table 4 Goodness-of-fit measures for pedestrian crash-frequency models | | | |  |  |
| Base model 0.25 mile | | | | 0.50 mile | 1 mile |

13 Note: LLN denotes the log likelihood at convergence for Base + NB model.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model statistics | NB | RPNB | NB | RPNB | NB | RPNB | NB | RPNB |
| Number of observers | 279 | 279 | 279 | 279 | 279 | 279 | 279 | 279 |
| Number of parameters | 4 | 5 | 7 | 8 | 7 | 8 | 7 | 8 |
| Log likelihood at convergence | -399.51 | -397.89 | -386.34 | -386.26 | -387.27 | -387.16 | -388.52 | -388.13 |
| AIC | 807.03 | 805.77 | 786.68 | 788.51 | 788.54 | 790.32 | 791.05 | 792.27 |
| Log-likelihood ratio test | | | | | | | | |
| 2 = -2(LLN-LLA) | 3.256 | | 26.352 | 26.515 | 24.487 | 24.712 | 21.983 | 22.761 |
| Degrees of freedom | 1 | | 3 | 4 | 3 | 4 | 3 | 4 |
| P-value | 0.08 | | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| Explanatory power of macroscopic factors | | | | | | | | |
| Variance of the error term | 0.785 | 0.613 0.585 0.578 | | | 0.595 | 0.587 | 0.601 | 0.595 |
| Proportion of reduction in variance, PRV |  | 21.96% 25.57% 26.37% | | | 24.26% | 25.21% | 23.48% | 24.18% |

# 14

15

16

1. *4.2 Parameter estimates and marginal effects*
2. Three types of crash-frequency models for motor vehicles, bicycles and pedestrians were
3. estimated and each type involved eight separate models based on four model specifications and
4. two analytical methods, yielding a total of 24 models. Three full models which have a better
5. goodness-of-fit, including motor vehicle NB model with 0.25 mile buffer width macroscopic
6. variables, bicycle NB model with 0.25 mile buffer width macroscopic variables and pedestrian NB
7. model with 0.25 mile buffer width macroscopic variables, were selected as recommended models
8. for reasons outlined. Meanwhile, the parameter estimates of three base NB models for motor
9. vehicles, bicycles and pedestrians were presented for comparison. Tables 5-7 show the parameter
10. estimates and their t-statistics for motor vehicle, bicycles and pedestrians crash-frequency models,
11. respectively. Since the model has nonlinear coefficients, direct parameters will not show a unit
12. effect on the number of crashes. Table 8 thus summarizes the results for marginal effects, which
13. could be interpreted as the average impact of a unit change in an explanatory variable on crash
14. frequency.
15. Comparing marginal effects of microscopic variables (traffic volume and road variables)
16. between the full NB models and the base NB models revealed some important differences. The
17. major differences were in marginal estimates of the variable ‘presence of traffic signal.’ The
18. results in the base model using NB approach showed that ‘presence of traffic signal’ had positive
19. association with the crash frequency for all three transportation modes; while this variable
20. became no statistically significant for motor vehicle and bicycle crashes in the full NB models.
21. Meanwhile, the present of macroscopic variables also modified the marginal effects of
22. microscopic variables (see Table 8). For example, the marginal effects of ln(AADT-minor) were
23. 6.38 and 0.14 respectively for motor vehicle and bicycle crashes in the base models, while these
24. values were 7.67 for motor vehicle crashes and 0.17 for bicycle crashes in models with
25. macroscopic variables. This difference clearly showed that in the absence of important
26. macroscopic variables, the marginal effects of ln(AADT-minor) for motor vehicle and bicycle
27. crashes were biased downwards by 16.8% and 17.6% respectively. These results agreed with the
28. safety research by Mitra and Washington (2012) that the exclusion of important variables may
29. cause bias in coefficient estimates and incorrect inferences.
30. According to the results of parameter estimates (Tables 5-7) and their marginal effects (Table
31. 8), significant variable sets for crashes were not consistent for different transportation modes.
32. ‘AADT on major approach’ and ‘density of total population’ were two contributing factors that
33. had statistically significant effects on the three response variables (i.e., motor vehicle, bicycle and
34. pedestrian crash frequency). Four variables including ’AADT on minor approach,’ ‘number of
35. legs’, ‘proportion of college productions and attractions’ and ‘proportion of workers commuting
36. by public transportation’ were significant for two response variables. Six variables were solely
37. associated with one response variable: ‘presence of traffic signal’, ‘speed limit on major
38. approach,’ ‘speed limit on minor approach,’ ‘proportion of home-based productions and
39. attractions,’ ‘proportion of population between age 16 and 64’ and ‘proportion of workers
40. commuting by walking’. The detailed interpretations for these significant risk factors are offered
41. in the following.

# 42

43

44

45

46

47

1. Table 5 Parameter estimates for motor vehicle crash-frequency models

Base NB model Full NB model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Mean | Standard Error | t-Statistic | Mean | Standard Error | t-Statistic |
| Ln(AADT-major) | 0.728 | 0.038 | 19.13 | 0.728 | 0.042 | 17.30 |
| Ln(AADT-minor) | 0.097 | 0.041 | 2.35 | 0.117 | 0.043 | 2.74 |
| Leg-number | 0.451 | 0.066 | 6.83 | 0.414 | 0.068 | 6.11 |
| traffic signal | 0.159 | 0.058 | 2.75 |  |  |  |
| Speed-Minor | 0 .031 | 0.004 | 8.30 | 0.028 | 0.006 | 4.49 |
| HB\_prop |  |  |  | -2.015 | 0.513 | -3.93 |
| Col\_prop |  |  |  | -2.324 | 0.673 | -2.64 |
| Pop\_density |  |  |  | 0.039 | 0.015 | 2.70 |
| Pub\_prop |  |  |  | -0.761 | 0.349 | -2.14 |
| Intercept |  |  |  | 1.377 | 0.467 | 2.95 |

1. Note: all parameters are significant at the 0.1 level or better.

# 3

1. Table 6 Parameter estimates for bicycle crash-frequency models

Base NB model Full NB model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Mean | Standard Error | t-Statistic | Mean | Standard Error | t-Statistic |
| Ln(AADT-major) | 0.490 | 0.130 | 3.77 | 0.580 | 0.127 | 4.58 |
| Ln(AADT-minor) | 0.132 | 0.085 | 1.55 | 0.165 | 0.086 | 1.92 |
| Leg-number | 0.369 | 0.171 | 2.15 | 0.464 | 0.173 | 2.69 |
| traffic signal | 0.312 | 0.135 | 2.32 |  |  |  |
| Col\_prop |  |  |  | -2.791 | 1.669 | -1.67 |
| Pop\_density |  |  |  | 0.082 | 0.031 | 2.64 |
| Wal\_ prop |  |  |  | 10.638 | 3.782 | 2.81 |
| intercept | -2.304 | 0.412 | -5.59 | -3.283 | 0.447 | -7.35 |

1. Note: all parameters are significant at the 0.1 level or better.

# 6

1. Table 7 Parameter estimates for pedestrian crash-frequency models

Base NB model Full NB model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Mean | Standard Error | t-Statistic | Mean | Standard Error | t-Statistic |
| Ln(AADT-major) | 0.877 | 0.161 | 5.45 | 0.924 | 0.162 | 5.69 |
| traffic signal | 0.732 | 0.154 | 4.75 | 0.522 | 0.156 | 3.35 |
| Speed-Major | -0.076 | 0.013 | -5.91 | -0.049 | 0.018 | -2.76 |
| Pop\_density |  |  |  | 0.090 | 0.031 | 2.88 |
| Age 16 to 64\_prop |  |  |  | -3.089 | 1.112 | -2.78 |
| Wal\_ prop |  |  |  | 11.211 | 3.809 | 2.94 |

1. Note: all parameters are significant at the 0.1 level or better.

# 9

10

11

1. Table 8 Estimate results for marginal effects of risk factors

Motor vehicle Bicycle Pedestrian

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Base | Full | Base | Full | Base | Full |
| Ln(AADT-major) | 48.00 | 47.65 | 0.50 | 0.59 | 1.15 | 1.19 |
| Ln(AADT-minor) | 6.38 | 7.67 | 0.14 | 0.17 |  |  |
| Leg-number | 26.60 | 24.48 | 0.35 | 0.43 |  |  |
| traffic signal | 10.47 |  | 0.32 |  | 0.96 | 0.67 |
| Speed-Major |  |  |  |  | -0.10 | -0.06 |
| Speed-Minor | 2.03 | 1.87 |  |  |  |  |
| HB\_prop |  | -131.94 |  |  |  |  |
| Col\_prop |  | -152.21 |  | -2.85 |  |  |
| Pop\_density |  | 2.58 |  | 0.08 |  | 0.12 |
| Age 16 to 64\_prop |  |  |  |  |  | -3.96 |
| Pub\_prop  Wal\_ prop |  | -49.86 |  | 10.88 |  | 14.39 |

1. Note: all parameters are significant at the 0.1 level or better.
2. *4.2.1 Traffic volume*
3. Similar to numerous prior studies, traffic volumes are significant variables for intersection
4. crashes and are positively correlated with crash occurrence (Lee and Abdel-Aty, 2005; Mitra and
5. Washington, 2012; Xie et al., 2013). The marginal effects of ln(AADT-major) were 47.65 , 0.59 and
6. 1.19 respectively for motor vehicle, bicycle and pedestrian crashes in the full models, indicating
7. that an average of thousand increase in major approach AADT will lead to a 2.31, 0.03 and 0.06
8. increase in motor vehicle ,bicycle and pedestrian crash frequency respectively. Similarly, an
9. average of thousand increase in minor approach AADT was associated with a 0.69 and 0.02
10. increase in motor vehicle and bicycle crashes.
11. *4.2.2 Number of legs*
12. The marginal effects of ‘number of legs’ on motor vehicle and bicycle crashes were 24.28 and
13. 0.43. This result suggested that the four-legged intersection was associated with 24.28 more motor
14. vehicle crashes and 0.43 more bicycle crashes compared to the intersection with three legs. This
15. result was generally expected and agreed with the preliminary finding that a larger number of
16. legs may increase the likelihood of crash occurrence due to more potential conflicts (Zeng and
17. Huang, 2014). However, this variable did not found to have significant effects on pedestrian crash
18. frequency. This may be due to the higher design standards of facilities (such as marked
19. crosswalks) in the large-leg intersection that leads to mixed effects of ‘number of legs’ on
20. pedestrian safety.
21. *4.2.3 Traffic signal*
22. The effects of traffic signal on safety are very interesting. The results of parameter estimates
23. in the base model using NB approach showed that ‘presence of traffic signal’ had positive
24. association with all three target variables. This is not consistent with the empirical hypothesis that
25. the installation of traffic lights could improve intersection safety. In the base model using RPNB
26. approach, ‘presence of traffic signal’ had a positive effect on three target variables but with a
27. varying magnitude across intersections (the results for the RPNB model were not presented since
28. there was no significant improvement in goodness-of-fit compared to the NB model). More
29. interestingly, this variable became no statistically significant for both motor vehicle and bicycle
30. crashes in the models integrating macroscopic variables. The possible reason for this difference is
31. that the macroscopic variables could account for a positive and heterogeneity effect of the
32. ‘presence of traffic signal’. This implies that the installation of traffic lights itself will not increase
33. the crash risk but traffic lights are always installed at relatively hazardous sites; however, more
34. work is needed to verify this conclusion.
35. *4.2.4 Speed limit*
36. The variables related to speed limit had inconsistent effects on crash occurrence by different
37. modes. ‘Speed limit on minor approach’ was positively correlated with the frequency of motor
38. vehicle crashes. An increase of 10 mph in speed limit on minor approach will increase motor
39. vehicle crashes by 18.7. This is generally expected since at high speeds the time to react to
40. changes in the environment is shorter, leading to higher crash frequency. However, ‘speed limit
41. on major approach’ was negatively associated with pedestrian crashes. The frequency of
42. pedestrian crashes will decrease by 0.6 with per 10 mph increase in speed limit on major
43. approach. The probable reason for this result is that a higher speed limit is always related with
44. higher design standards of facilities such as pedestrian overcrossing and underpass. The effect of
45. speed limit on bicycle crashes was not significant.
46. *4.2.5 Trip characteristics*
47. Results showed that the proportion of home-based productions and attractions near an
48. intersection was negatively associated with motor vehicle crashes. The frequency of motor
49. vehicle crashes will decrease by 1.32 with a percentage increase in proportion of home-based
50. trips. This is reasonable since drivers of a home-based trip are more familiar with the traffic
51. environment and have more cautious driving behaviors (Abdel-Aty et al., 2011). ‘Proportion of
52. college productions and attractions’ was negatively associated with motor vehicle and bicycle
53. crashes while was not statistically significant for pedestrian crashes. A percentage increase in
54. proportion of college trips will result in a 1.52 and 0.03 decrease in motor vehicle and bicycle
55. crashes. The decreased motor vehicle and bicycle crashes may be due to better traffic control
56. measures in these areas. However, higher proportion of college productions and attractions,
57. which is always related to better traffic control measures and high number of walking trips, leads
58. to mixed effects on pedestrian crashes.
59. *4.2.6 Demographic characteristics*
60. ‘Density of total population’ was the influential macroscopic variable and was positively
61. associated with all three target variables. The marginal effects of population density near an
62. intersection showed that an increase in population density will increase the frequency of motor
63. vehicle, bicycle and pedestrian crashes by 2.48, 0.08, and 0.12. This agrees with previous studies
64. as a larger population is always consistent with more opportunities in terms of crash exposure
65. (Lee et al., 2015; Mitra and Washington, 2012; Pulugurtha and Sambhara, 2011). In addition, the
66. proportion of population between age 16 and 64 near an intersection was found to have negative
67. effects on pedestrian crashes. The frequency of pedestrian crashes will decrease by 0.04 with a
68. percentage increase in proportion of population aged 16-64. This may be due to middle-aged
69. people walking less in comparison to young and/or old people, as well as having better ability in
70. avoiding crash risk (Huang et al., 2010).
71. *4.2.7 Commute behaviors*
72. A percentage increase in proportion workers commuting by public transportation near an
73. intersection was associated with a 0.50 decrease in motor vehicles and a 0.10 increase in bicycle
74. crashes, indicating that the public transportation had opposite effects on motor vehicles and
75. bicycle crashes. In addition, ‘proportion of workers commuting by walking’ had significant and
76. positive associations with pedestrian crash occurrence. The number of pedestrian crashes will
77. increase by 0.14, with a percentage increase in proportion workers commuting by walking. This
78. result is not surprising since walking is always associated with the exposure of pedestrian
79. crashes.

## 5. Conclusions and recommendations

1. This paper sought to examine the effects of omitted macroscopic factors in crash-frequency
2. models by transportation modes at intersections. For this purpose, several separate
3. intersection-level crash-frequency model for motor vehicle, bicycle and pedestrian modes were
4. developed. Road characteristics related to geometric design and regulatory/control attributes and
5. traffic characteristics of the intersection entities, as well as macroscopic factors including trip
6. production/attraction, demographic and socio-economic characteristics and commute behaviors
7. at TAZ level surrounding the intersection, were used as explanatory variables. Those data
8. extracted for 279 intersections located in Hillsborough County, Florida, USA, were used for
9. model development.
10. The empirical analysis revealed a number of interesting findings. First, omission of
11. macroscopic variables would result in biased estimation of retained microscopic variables.
12. Results for marginal effects of traffic volumes and road features showed significant differences
13. between in the base model and the full model with macroscopic variables. For example, the safety
14. effect of minor approach AADT on motor vehicle and bicycle crashes are biased downwards by
15. 16.8% and 17.6% in the absent of macroscopic variables.
16. Second, macroscopic variables had potential effects in tracking the heterogeneity of certain
17. risk factors. The results in the base model using RPNB approach showed that the safety effect of
18. ‘presence of traffic signal’ was best fit with a normally distributed random parameter suggesting
19. ‘presence of traffic signal’ had heterogeneous effects across intersections; while this variable
20. became no statistically significant in models with macroscopic variables. This implied that the
21. heterogeneous effects of ‘presence of traffic signal’ on motor vehicles crashes could be mostly
22. captured by these macroscopic variables, at least for the Hillsborough dataset.
23. Third, model comparison using log likelihood at convergence suggested that considering
24. macroscopic variables was vital in elevating the model performance. In addition, the values of
25. PRV were further calculated to assess the explanatory power of macroscopic variables. Results
26. showed the values of PRV in bicycle and pedestrian crash-frequency model were much higher
27. than in motor vehicle models, indicating that integrating macroscopic factors played a more
28. important role in developing non-motorized crash-frequency model than in developing motor
29. vehicle models.
30. Fourth, comparing model outputs developed based on 0.25 mile, 0.5 mile and 1 mile buffer
31. width data, models with macroscopic variables of 0.25 mile buffer width had the lowest AIC and
32. highest PRV, conversely, the models with macroscopic variables of 1 mile buffer width had the
33. highest AIC and lowest PRV in all three types of crash-frequency models by modes. Thus a
34. relatively smaller buffer width to extract macroscopic factors around the intersection would
35. provide a better estimate.
36. Finally, macroscopic factors of the surrounding zone of an intersection, such as ‘proportion
37. of home-based productions and attractions’, ‘proportion of college productions and attractions’,
38. ‘density of total population’, ‘proportion of population between age 16 and 64’, proportion of
39. workers commuting by public transportation and ‘proportion of workers commuting by walking’,
40. were demonstrated to have significant effects on intersection crashes; while these variables are
41. always ignored in traditionally micro-level (e.g., intersections and segments) crash frequency
42. model. This indicated that not only traffic volumes and road features but also macroscopic factors
43. should be considered in estimating crash risk and identifying crash-prone locations.
44. The topic of integrating macroscopic factors in intersection/segment-level crash-frequency
    1. model is emerging. This study has great research potential in pro-actively predicting crash risk
    2. and identifying suitable countermeasures to reduce the crashes at “new” intersections/segments
    3. as well as intersections/segments near “new” development. Nevertheless, several limitations
    4. should be noted for this study. First, it is worthwhile to apply this model to other intersections
    5. and regions in order to investigate spatial transferability of the calibrated models. In addition,
    6. there may be some correlations among crash frequency by different transportation modes within
    7. intersections, which is caused by some unobserved influential factors. Therefore, a simultaneous
    8. model accounting for correlations of crashes among transportation modes, such as the
    9. multivariate model, will be further explored to investigate this issue.

## Acknowledgements

* 1. This work was jointly supported by: 1) Natural Science Foundation of China (No.71371192,
  2. No. 71561167001); 2) the Research Fund for Fok Ying Tong Education Foundation of Hong Kong
  3. (142005); and 3) Fundamental Research Funds for the Central Universities of CSU (No.
  4. 2016zzts050). We would like to thank Dr. Mohamed Abdel-Aty at the University of Central
  5. Florida and the Florida Department of Transportation for providing the data.

## References

* 1. Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway
  2. characteristics to manage safety in traffic analysis zones. Transportation Research Record:
  3. Journal of the Transportation Research Board 2213, 20-28.
  4. Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies
  5. with random-parameters count models. Accident Analysis and Prevention 41, 153-159.
  6. Christoffel, T., Gallagher, S., 1999. Injury Prevention and Public Health: Practical Knowledge,
  7. Skills, and Strategies. Aspen Publishers Inc., Gaithersburg, MD.
  8. Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of
  9. traffic analysis zones: Assessing the spatial proximity effects. Accident Analysis and

26 Prevention 82, 192–198.

1. Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: A
2. Bayesian spatio-temporal interaction approach. Accident Analysis and Prevention 92,

29 256-264.

1. Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida:
2. Bayesian spatial modeling. Transportation Research Record: Journal of the Transportation
3. Research Board 2148, 27-37.
4. Huang H., Song B., Xu P., Zeng Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal
5. crash prediction with application in hot zones identification. Journal of Transport Geography,

35 54, 248-256.

1. Jacobsen, P.L., 2003. Safety in numbers: More walkers and bicyclists, safer walking and bicycling.
2. Injury Prevention 9, 205-209.
3. Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at
4. intersections in Florida. Accident Analysis and Prevention 37, 775-786.
5. Lee, J., Abdel-Aty, M., Jiang X., 2015. Multivariate crash modeling for motor vehicle and
6. non-motorized modes at the macroscopic level. Accident Analysis and Prevention 78,

42 146-154.

1. Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and
2. assessment of methodological alternatives. Transportation Research Part A 44, 291-305.
3. Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash
4. modeling. Accident Analysis and Prevention 49, 439-448.
   1. Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment,
   2. pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections.
   3. Accident Analysis and Prevention 43, 1624-1634.
   4. Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier
   5. and future directions. Analytic Methods in Accident Research 1, 1-22.
   6. Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical
   7. analysis of highway accident data. Analytic Methods in Accident Research 11, 1-16.
   8. NHTSA, National Highway Traffic Safety Administration, FARS Data, 2013.
   9. www-fars.nhtsa.dot.gov/Main/index.aspx
   10. Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized
   11. intersections. Accident Analysis and Prevention 43, 439-446.
   12. Quddus, M., 2008. Modeling area-wide count outcomes with spatial correlation and
   13. heterogeneity: an analysis of London crash data. Accident Analysis and Prevention 40,

14 1486-1497.

1. Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical linear models: Applications and data analysis
2. methods, 2nd ed. Sage, Thousand Oaks.
3. Ryb, G.E., Dischinger, P.C., Kufera, J.A., Soderstrom, C.A., 2007. Social, behavioral and driving
4. characteristics of injured pedestrians: A comparison with other unintentional trauma patients.
5. Accident Analysis and Prevention 39, 313-318.
6. Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at
7. signalized intersections: a Bayesian modeling approach. Accident Analysis and Prevention 59,

22 9-17.

1. Strauss, J., Miranda-Moreno, L.F., Morency, P., 2014. Multimodal injury risk analysis of road users
2. at signalized and non-signalized intersections. Accident Analysis and Prevention 71, 201-209.
3. Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai G., Isa-Tavarez, J., 2012. The role of built
4. environment on pedestrian crash frequency. Safety science 50, 1141-1151.
5. Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate
6. crash frequencies by severity, number of vehicles involved, collision and location type.
7. Accident Analysis and Prevention 59, 309-318.
8. Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint
9. model. Accident Analysis and Prevention 90, 152-158.
10. Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis
11. in Shanghai, China using Bayesian hierarchical models. Accident Analysis and Prevention 50,

34 25-33.

1. Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional
2. safety modeling: Identifying and assessing the modifiable areal unit problem. Accident
3. Analysis and Prevention 70, 110–120.
4. Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus
5. geographically weighting. Accident Analysis and Prevention 75, 16-25.
6. Xu, P., Huang, H., Dong, N., Wong, S., 2016. Revisiting crash spatial heterogeneity: a Bayesian
7. spatially varying coefficients approach. Accident Analysis and Prevention, DOI:

42 10.1016/j.aap.2016.10.015.

1. Zeng, Q., Huang, H., 2014. Bayesian spatial joint modeling of traffic crashes on an urban road
2. network. Accident Analysis and Prevention 67, 105-112.

17

[View publication stats](https://www.researchgate.net/publication/309308779)